The Mathematics of Evolution

Steven R. Lay, Department of Natural Sciences and Mathematics,
Lee University, Cleveland, TN, 37311, USA, slay@leeuniversity.edu

Abstract

Like many universities, Lee University has a non-major's mathematics course for liberal arts students. The course typically includes a potpourri of topics: logical thinking, scientific notation, linear functions, estimation, and probability. At Lee, we have found a way to conclude the course that applies these varied topics to an issue designed to engage student interest and promote critical thinking. We have developed a series of three lessons on the mathematics of evolution. This paper includes a sampling of the topics included in those lessons.

One of the fundamental questions of life is **Where did life come from?** Many people believe that man is the product of a random natural process called **evolution**. Millions of years ago, "simple" life forms appeared on the earth. Over time, these simple forms gradually became more complex. Eventually, man appeared.

What does mathematics say about the hypotheses and equations that provide the scientific background for the theory of evolution?

Radiometric Dating

Since evolution is dependent on the earth being millions of years old, we begin by asking, "How old is the earth?" Estimates of the age of the earth vary widely from 6,000 years to 4.5 billion years. Those who hold to the recent creation of the earth are sometimes called "young-earth" creationists and those who subscribe to a more ancient origin are said to believe in an "old-earth." There are highly trained scientists (and devout Christians) in both groups. Most attempts to date the age of the earth are based on properties of radioactive decay. For example, fossils can be dated with Carbon-14 and igneous rocks with Potassium-40. These techniques **assume** that the rate of decay has been constant through all time. Indeed, over the last hundred years, the rates have appeared to be constant. But can we conclude that they have always been the same?

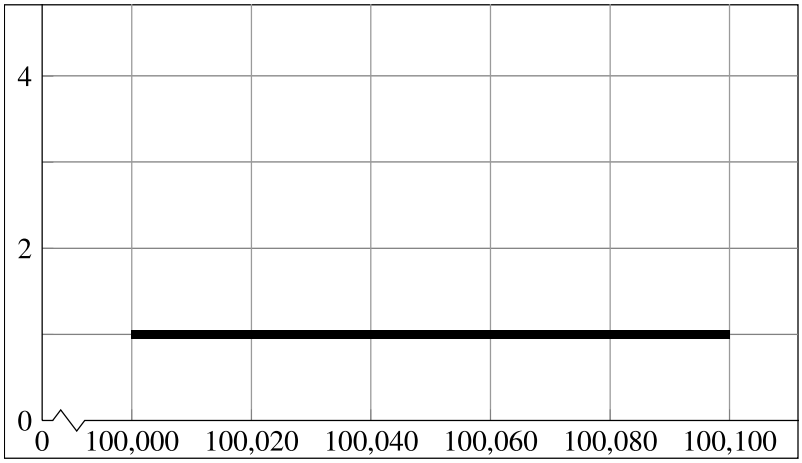
Extrapolation: Using the Present to Measure the Past

Suppose for the moment that the earth is 100,000 years old. Let t represent time in years since the formation of the earth and suppose we measure some physical quantity $Q(t)$ that may or may not change with time. For example, $Q(t)$ might be the ratio of the half-life of ^{14}C at time t to the half-life of ^{14}C measured in the year 1900 A.D. If $Q(t)$ is constant, then the half-life of ^{14}C would be constant, too. We consider three possible functions that might model the behavior of Q over time.[1]

$$f(t) = \frac{t^2 + t}{t^2 + 1}, \quad g(t) = 1, \quad \text{and} \quad h(t) = \frac{t^2 + 1}{t^2 + t}.$$

Which of these three functions do you think is graphed below, where we have shown the portion of the graph from $t = 100,000$ to $t = 100,100$?

[1]For the sake of simplicity, we have chosen functions that have values near 1. By taking multiples of these functions, we could make them approach any finite value.



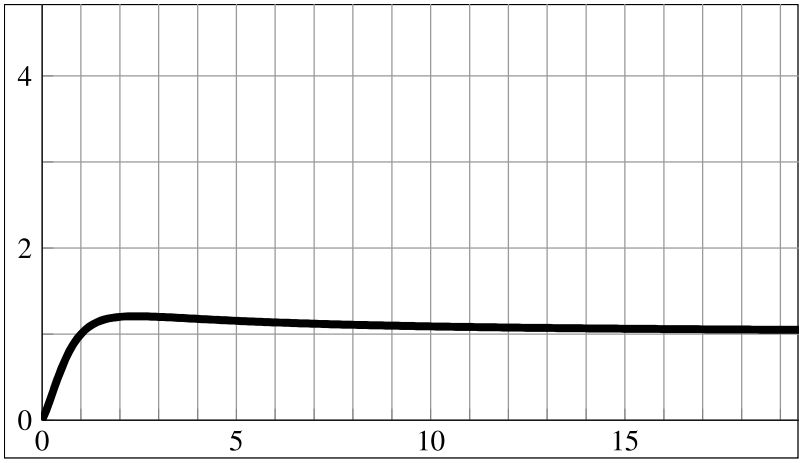
Graph for $t = 100,000$ to $t = 100,100$ Years

Actually, the graph could be any one of these three functions since they are indistinguishable over the range of $t = 100,000$ to $t = 100,100$ as shown. Here is a table of the values for the three functions, where we have truncated the numbers to nine decimal digits:

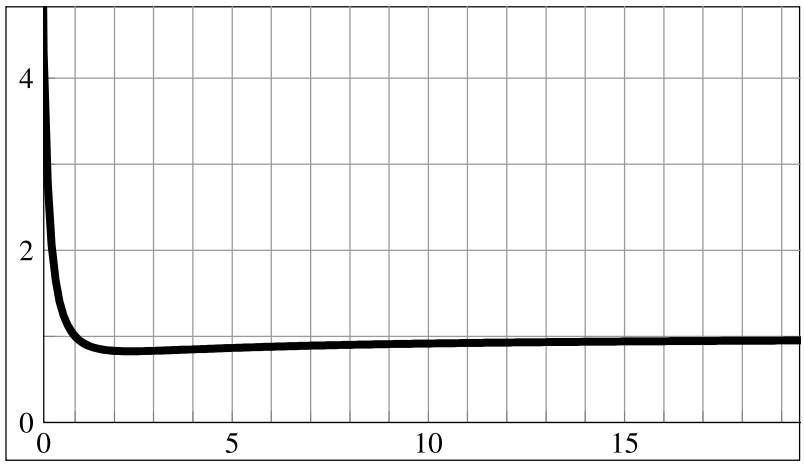
Time in years	$f(t) = \frac{t^2+t}{t^2+1}$	$g(t) = 1$	$h(t) = \frac{t^2+1}{t^2+t}$
$t = 100,000$	$1,000\,009\,999$	$1.000\,000\,000$	$0.999\,990\,000$
$t = 100,050$	$1,000\,009\,994$	$1.000\,000\,000$	$0.999\,990\,005$
$t = 100,100$	$1,000\,009\,989$	$1.000\,000\,000$	$0.999\,990\,010$

As we progress through the 100 years, the function $f(t)$ is a tiny bit more than 1 and decreasing just slightly. The function $g(t)$ is constant at 1. The function $h(t)$ is a tiny bit less than 1 and increasing slightly. We observe that all three functions are virtually constant and their differences are so small they cannot be seen. But what happens to the functions $f(t)$ and $h(t)$ for values of t close to zero? Are the functions "constant" there?

Here are the graphs of the functions near $t = 0$:



Graph of $f(t) = \frac{t^2+t}{t^2+1}$ near $t = 0$



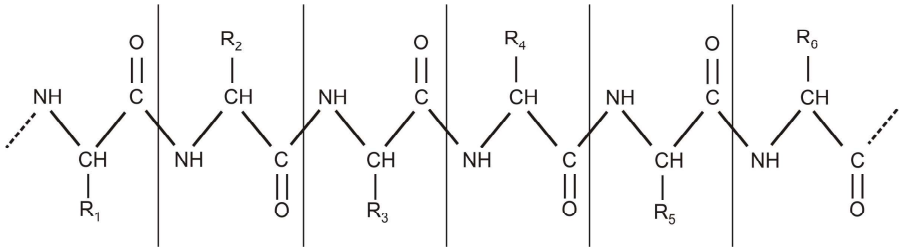
Graph of $h(t) = \frac{t^2+1}{t^2+t}$ near $t = 0$

As we go from right to left, the f function increases slightly and then turns and falls abruptly to 0. The h function decreases slightly to a minimum point and then rises without bound. So what appears to be constant in the present is far from constant in the past.

These graphs illustrate the danger of taking an apparent trend that appears in a relatively short period of time and extrapolating that trend over a great many years. Since we cannot know the initial conditions at time $t = 0$, there is no way of knowing for certain that what appears to be constant now has always been constant in the past. And without that knowledge, there is no way to be sure of the age of the earth.

The Origin and Progression of Life

Proteins are the basic building blocks of life. They consist of long strings of amino acids.



There are 20 kinds of amino acids that combine together in a specific order for each protein. Some proteins contain more than 20,000 amino acids, but the probability of even a small protein developing by random chance is so small it is hard to distinguish it from zero. For example, pancreatic ribonuclease is a small protein made up of a string of 127 amino acids. Could it have evolved by means of a sequence of random mutations in the DNA that controls the production of protein in the cell?

Almost all mutations are either neutral or harmful to an organism's ability to reproduce [4]. But let's be generous and suppose that 1 out of every 10 mutations to the DNA in a cell is beneficial in two ways:

- It moves the DNA one step closer to enabling the cell to produce pancreatic ribonuclease. That is, the protein synthesized by the DNA has one more amino acid in agreement with pancreatic ribonuclease.

- It alters the DNA in such a way that the cell is not destroyed by natural selection.

Since pancreatic ribonuclease requires the exact ordering of 127 amino acids, its production could be accomplished by a sequence of 127 beneficial mutations, each of probability $1/10$. This gives a combined probability of $(1/10)^{127} = 10^{-127}$ or 1 out of every 10^{127} trials. Let's see how long this might take, given a most favorable environment.

Example. Produce a Simple Protein by Mutations
Suppose that the 5.5×10^{15} *square feet of the earth's surface is covered entirely by mutating cells with one billion* (10^9) *cells per square foot. And suppose that each cell undergoes a random mutation at the rate of one mutation every second, with* $1/10$ *of those mutations being beneficial (as described above). How long would we expect it to take for one of the cells to develop the ability to synthesize pancreatic ribonuclease?*

We need on average to have 10^{127} trials to generate one cell capable of synthesizing pancreatic ribonuclease. Since it requires 127 mutations for each trial and mutations occur at one per second, it will take 127 seconds for a cell to go through its 127 mutations. Then, if it is unsuccessful, it can start over again with a new trial.[2] Thus, worldwide there are

$$(5.5 \times 10^{15})(10^9) = 5.5 \times 10^{24} \text{ trials every } 127 \text{ seconds}$$

or, on average, $\frac{5.5 \times 10^{24}}{127} \approx 4.33 \times 10^{22}$ trials a second. Since we need 10^{127} trials, this will take

$$\frac{10^{127}}{4.33 \times 10^{22}} = \left(\frac{10}{4.33}\right)\left(\frac{10^{126}}{10^{22}}\right) \approx 2.31 \times 10^{104} \text{ seconds.}$$

Since there are 3.16×10^7 seconds in a year, we would expect it to take

$$\frac{2.31 \times 10^{104}}{3.16 \times 10^7} = \left(\frac{23.1}{3.16}\right)\left(\frac{10^{103}}{10^7}\right) \approx 7.31 \times 10^{96} \text{ years.}$$

This is trillions of trillions times the maximum age of the universe: 14×10^9 or 14 billion years.

Of course, pancreatic ribonuclease is of little benefit to the creature that has randomly generated the enzyme unless a fully functioning pancreas (which is far more complex than the enzyme) emerges simultaneously. We pursue this further in the next section which looks more closely at how the DNA code works.

The Genetic Code

Every human life begins as a single cell: a fertilized egg. This first cell divides into two, the two into four, and so on. Soon the cells begin to differentiate: some into heart muscle, some into arms and fingers, and some into eyes or ears. Over the next few months, hundreds of bodily systems are constructed in a precise way at a precise time. And all the information necessary to direct the construction and operation of these systems is somehow coded in the DNA of that first cell. How is that possible? How can so much information be compressed into such a small space?

Part of the answer to that question is found in the design of the DNA code itself. DNA consists of long twisted strands of four nucleotides abbreviated **A**, **T**, **C**, and **G**. The primary DNA code groups these nucleotides into triplets called codons. There are 64 ($4^3 = 64$) possible codons, and each one encodes for one of the 20 amino acids used in the synthesis of proteins. It turns out that most sequences of DNA are poly-functional. That

[2]To simplify the computations, we assume that whenever a cell is destroyed by natural selection, it is replaced by a new cell in the next set of trials so that the population remains constant at one billion cells per square foot.

is, they can encrypt multiple overlapping codes.[3] For example, the codon **CTG** encodes for the amino acid Leucine. But when read backwards (**GTC**) it encodes for Valine.

$\longrightarrow$ Leucine

CTG

Valine $\longleftarrow$

If the **CTG** codon is followed by the **CCG** codon (Proline) and the first nucleotide is skipped, then the first codon would become **TGC** (Cysteine).

~~C~~**TGCCG** $\longrightarrow$ Cysteine

To illustrate the power and weakness of a poly-functional code, we look at an example of a mathematical code that has five levels of meaning. Suppose we take as our encrypted message (M) the first 16 digits in the decimal expansion of π:

$$M = \ 3\ 1\ 4\ 1\ 5\ 9\ 2\ 6\ 5\ 3\ 5\ 8\ 9\ 7\ 9\ 3$$

and we use Translation Table 1.

	0	1	2	3	4	5	6	7	8	9
0	B	□	S	□	S	T	□	□	□	A
10	K	M	S	D	O	W	O	I	I	R
20	N	□	L	T	D	V	I	E	H	E
30	R	□	A	S	X	□	□	L	I	C
40	V	S	F	□	F	□	Y	N	E	K
50	A	Q	O	V	E	E	N	G	E	L
60	M	I	W	□	□	Z	U	S	A	I
70	□	H	G	D	A	I	J	O	E	U
80	E	D	L	S	H	T	U	□	Y	G
90	R	□	P	□	C	O	S	R	N	E

Table 1

When we write the digits in M as pairs of numbers we get the following translation for level one:

(Level 1)

31	41	59	26	53	58	97	93
□	S	L	I	V	E	R	□

If we reverse the digits in M and group them in pairs, we get the Level 2 message:

[3] Some of the additional codes are based on reading the codons backwards or starting the translation at a different nucleotide, as illustrated here with Leucine, Valine, and Cysteine. Recent research has found other codes that are not based on consecutive codons of nucleotides. (See [7] and [8].)

(Level 2)

39	79	85	35	62	95	14	13
C	U	T	□	W	O	O	D

To get the Level 3 message, return to the original message M. Begin by dropping the first three digits and the last five digits:

1 5 9 2 6 5 3 5

Now replace each digit by a number pair that is the sum of the digits up to (and including) that digit. For example, the first pair is 1 (written as 01 so that there are two digits). The second pair is $1 + 5 = 06$. The third pair is $1 + 5 + 9 = 15$, etc.

(Level 3)

01	06	15	17	23	28	31	36
□	□	W	I	T	H	□	□

The coding for Level 4 is more complicated. Start with the numbers from Level 3:

(Level 3) 01 06 15 17 23 28 31 36

Separate the digits from 0 to 9 into two groups: Those digits that have a pointed / flat top (1, 4, 5, and 7) and those digits that have a curved top (0, 2, 3, 6, 8, and 9). Then permute the digits within each group as follows:

$$1 \longrightarrow 7 \longrightarrow 4 \longrightarrow 5 \longrightarrow 1 \quad \text{and} \quad 0 \longrightarrow 8 \longrightarrow 2 \longrightarrow 9 \longrightarrow 6 \longrightarrow 3 \longrightarrow 0$$

This generates the message for Level 4.

(Level 4)

01	06	15	17	23	28	31	36
↓↓	↓↓	↓↓	↓↓	↓↓	↓↓	↓↓	↓↓
87	83	71	74	90	92	07	03
□	S	H	A	R	P	□	□

To obtain the fifth level, begin with the original message M. Drop the first two digits and the last six digits:

4 1 5 9 2 6 5 3

Now replace each digit by a number pair that is twice the sum of the digits up to (and including) that digit. For example, the first pair is $(2)(4) = 08$. The second pair is $(2)(4+1) = 10$. The third pair is $(2)(4+1+5) = 20$, etc. If the number is over 99, subtract 100 and write what is left.

(Level 5)

08	10	20	38	42	54	64	70
□	K	N	I	F	E	□	□

The message is now complete. The first level says "sliver." And the next four levels describe how a sliver might be obtained: "cut wood with sharp knife."

Suppose that a "mutation" occurs in the original message M. For example, suppose the fourth digit is changed from a 1 to a 3. Call this message M'.

$$M' = \ 3\ 1\ 4\ 3\ 5\ 9\ 2\ 6\ 5\ 3\ 5\ 8\ 9\ 7\ 9\ 3$$

Now use the same translation table and the same instructions at each level to decode the mutated message. In Level 1, when we group the digits in pairs, the second pair changes from 41 to 43. This changes the "S" to a

blank. And “sliver” changes into “liver.”

(Level 1)	31	43	59	26	53	58	97	93
	□	□	L	I	V	E	R	□

For Level 2, reverse the original M' digits and pair them:

(Level 2)	39	79	85	35	62	95	34	13
	C	U	T	□	W	O	X	D

For Level 3, drop the first 3 and the last 5 digits from M':

3 5 9 2 6 5 3 5

Then replace each digit by the sum of the digits up to that point:

(Level 2)	03	08	17	19	25	30	33	38
	□	□	I	R	V	R	S	I

For Level 4, permute the digits from Level 3:

$$1 \longrightarrow 7 \longrightarrow 4 \longrightarrow 5 \longrightarrow 1 \quad \text{and} \quad 2 \longrightarrow 9 \longrightarrow 6 \longrightarrow 3 \longrightarrow 0 \longrightarrow 8 \longrightarrow 2$$

	03	08	17	19	25	30	33	38
	↓↓	↓↓	↓↓	↓↓	↓↓	↓↓	↓↓	↓↓
(Level 4)	80	82	74	76	91	08	00	02
	E	L	A	J	□	□	B	S

And finally, for Level 5, drop the first two and last six digits from M':

4 3 5 9 2 6 5 3

Then replace each digit by twice the sum of the digits up to that point:

(Level 5)	08	14	24	42	46	58	68	74
	□	O	D	F	Y	E	A	A

At the first level, the message has changed from “sliver” to “liver.” This potentially makes some sense and might be considered an increase in the level of information. A liver (the human organ) is certainly more complex than a sliver. But the message contained in the other four levels has become unintelligible:

CUT WOXD IRVRSIELAJ BS ODFYEAA

It certainly tells us nothing about how to make a liver.

While a poly-functional code is very powerful in that it can compress a lot of information into a small space, it is also very weak when it comes to allowing beneficial mutations. A potentially desirable change in one level will inevitably destroy information at the other levels. So the random creation of new organs and biological systems (evolution) cannot be accomplished by mutations and natural selection.

In fact, the mathematics suggests that instead of the biological world evolving upward, it is slowly devolving to a lower level. That is, the genetic clock is running backwards, not forwards. Indeed, this is what biological

scientists are now finding [1, 2, 3, 5, 6]. For example, Dr. James Crow of the Genetics Laboratory at the University of Wisconsin writes,

> Since most mutations, if they have any effect at all, are harmful, the overall impact of the mutation process must be deleterious. [2, p. 8380]

And after careful analysis he concludes that "the decrease in viability from mutation accumulation is some 1 or 2% per generation."[4]

If evolution cannot even preserve the genetic information we currently possess, it certainly could not have "created" this information in the first place! There may be philosophical reasons for holding to the validity of the theory of evolution, but from a mathematical perspective, the evidence points strongly in the opposite direction.

Note: For information on obtaining a digital file of the complete set of lessons, please contact the author.

References

[1] T. Bataillon. Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? *Heredity*, 84:497–501, 2000.

[2] James F. Crow. The high spontaneous mutation rate: is it a health risk? *Proc. of the National Academy of Sciences*, 94:8380–8386, August 1997.

[3] Adam Eyre-Walker and Peter D. Keightley. High genomic deleterious mutation rates in hominids. *Nature*, 397(6717):344–347, January 28, 1999.

[4] Adam Eyre-Walker and Peter D. Keightley. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8:610–618, 2007.

[5] K. Kevin Higgins and Michael Lynch. Metapopulation extinction caused by mutation accumulation. *Proc. of the National Academy of Science*, 98:2928–2933, 2001.

[6] Michael Lynch, John Conery, and Reinhard Burger. Mutational meltdowns in sexual populations. *Evolution*, 49:1067–1080, December 1995.

[7] E. N. Trifonov. Multiple codes of nucleotide sequences. *Bulletin of Mathematical Biology*, 51:417–432, 1989.

[8] E. N. Trifonov, Z. Volkovich, and Z. M. Frenkel. Multiple levels of meaning in DNA sequences, and one more. *Annals of the New York Academy of Sciences*, 35:35–8, September 2012.

[4]While Dr. Crow concludes that mutations are driving us backward, not forward, he is not a prophet of doom. He is optimistic that advances in technology (particularly in molecular biology) will enable future generations to overcome their increasing genetic weaknesses. (See [2, p. 8385].)